Multivariate Functional Data Clusterization by PCA in Sobolev Space Using Wavelets

Ruslan Suyundykov, Stéphane Puechmorel, Louis Ferré

Multivariate Functional Data Clusterization by PCA in Sobolev Space Using Wavelets

R. Suyundykov
ENAC Dept. MI LMA
7, Avenue Edouard Belin
31055 Toulouse FRANCE
email: ruslan@recherche.enac.fr

S. Puechmorel
ENAC Dept. MI LMA
7, Avenue Edouard Belin
31055 Toulouse FRANCE
email: puechmor@recherche.enac.fr

L. Ferre
Universite Toulouse Le Mirail
5, alles Antonio Machado
31058 Toulouse FRANCE
email: loferre@univ-tlse2.fr

***Abstract*—Nous presentons une methode basee sur la classification en espace fonctionnel appliquee aux trajectoires avion. La particularite des donnees est de se presenter comme des fonctions deux fois continuement differentiables par morceaux, l'information discriminante se situant des les sauts. L'utilisation d'une dcomposition en ondelettes permet une implementation simple en espace de Sobolev et permet de separer correctement des trajectoires presentant des points de branchement.**

***Abstract*—We investigate reducing the dimensionality of multivariate functional data by using principal component analysis on wavelet coefficients in Sobolev space in order to cluster them. Dimensionality reduction using principal component analysis on wavelet coefficients in Sobolev space is investigated. Example on aircraft landing trajectories in Toulouse airport "Blagnac" is given.**

Keywords : Apprentissage et classification, Ondelettes

I. Introduction

The original motivation for this work comes from the field of air traffic control. It is forecasted that in future years the growth of the number of flights will require a paradigm shift in the way aircraft are controlled. Major projects (SESAR in EU and NEXTGEN in the USA) have been launched in order to improve the capacity of airspace by a factor at least 3 over the current situation. An increasing level of automation and autonomy is expected, requiring safe and robust trajectory planning. In this context, the basic object in Air Traffic Management (ATM) will be the trajectory as a whole replacing the current radar plots. In order to cope with this new framework, it is needed to gain knowledge on the statistics of aircraft trajectories, first on the basis of nowodays trafic and then to forecast to the expected one. Among the needs, it is highly desirable to have a mean of clustering trajectories, mainly into major flows and exceptional ones. Some work has already been done on this subject, but using the classical radar plots information, thus given only path of maximum aircraft density : altough most results are coherent, some high density paths turn out to be unrealistic since not corresponding to any trajectory. The contribution presented here starts from the trajectory point of view and uses techniques from functional data clustering in order to present an acceptable tool for the future ATM system. Furthermore, an innovative approach based on wavelet expansions and sobolov space classification will be described.

II. Understanding the data

A. The flight dynamics model

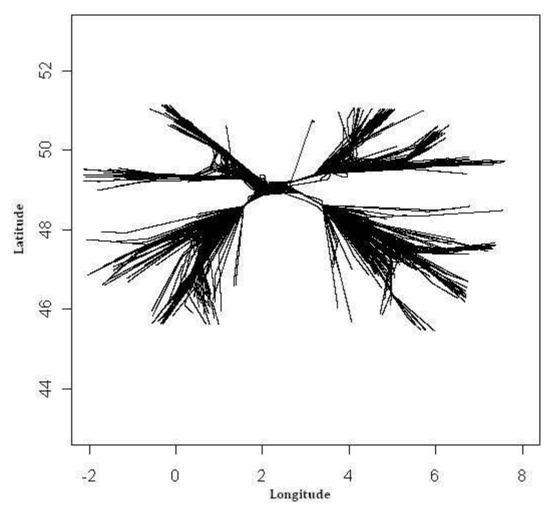

Figure 1. Example on aircraft landing trajectories in Toulouse airport "Blagnac" (longitude, latitude).

Aircaft are flying according to flight dynamics model that is in turn derived from Newton's second principle. The actions taken by the pilot can be modeled as torques or forces applied to the aircraft. As an example, engine thrust or airbrakes combined with longitudinal component of weight produce a force resulting in longitudinal acceleration, while actions on flaps induce a torque which in turn results in angular acceleration on the roll axis. Given the time scales involved in trajectory classification, one may assume that pilot actions are instantaneous and that the resulting force or torque consist of piecewise constant functions. Moreover, most of the time the aircraft is in cruise condition, namely no pilot actions are taking place. As a consequence, it is relevant for the application to make the following assumptions :

- The aircraft trajectories are C^1, piecewise C^2 functions.
- The external random perturbations on the trajectory are coming either from the wind (effect on velocity) or from measurement errors (effect on position).
- The discriminating factor between trajectories is the pilot actions.

As a consequence, the framework of choice will be functional principal component analysis applied to Sobolev space valued process.

B. Landing trajectories at Toulouse Blagnac Airport

We now consider aircraft landing trajectories in Toulouse airport "Blagnac" (see fig. 1). We can see obvious clusters corresponding to branches. The discriminating factor in such a case are the changes in velocity. By selecting only two branching bundles of trajectories within the database and computing the velocity, we can make this remark more obvious (see fig. 2, 3)

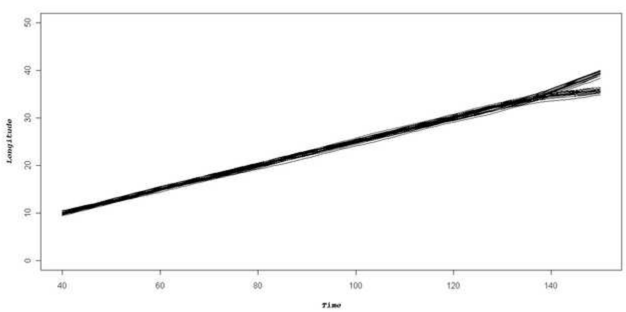

Figure 2. Example of aircraft trajectories longitude data.

Without loss of generality for clusterization we can suppose that all aircraft landing trajectories are defined on a fixed interval $\mathcal{T}$ and have landed at the same time. Each trajectory can be described as a sampled curve in $\mathbb{R}^3$ space. Formally, we have observations which consist of triple $\vec{f}^{(i)} = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)})^T$, where x_1, x_2, x_3 are functions on the interval $\mathcal{T}$. We assume that the curves from which the samples are obtained are C^1, piecewise C^2 (and thus belong to the Sobolev space $W^2(\mathbb{R})$, since $\mathcal{T}$ is a compact interval).

III. Wavelets

A. Introduction

The theory of wavelets was developed by Y.Meyer, I.Daubechies, S.Mallat and others in the end of 1980s (see [1],

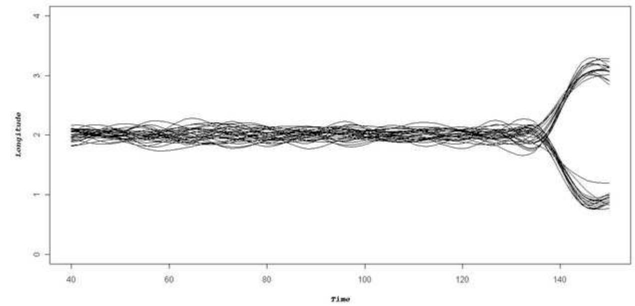

Figure 3. The estimated velocity of aircraft trajectories longitude data.

[3], [4]). Wavelets have proven to be a valuable alternative to traditional Fourier analysis when the content to be analyzed has singularities or fast variations. In our case, we are trying to locate fast changes in the velocity (first derivative of the trajectory) and we may expect that such a tool will have excellent clustering properties.

B. Wavelets in Sobolev space

Any $f \in L_2(\mathbb{R})$ can be represented as a series (convergent in $L_2(\mathbb{R})$):

$$f(t) = \sum_{k\in\mathbb{Z}} c_k \varphi_k(t) + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} c_{jk}\psi_{jk}(t), \qquad (2.1)$$

$$\varphi_k(t) = \varphi(t-k), k \in \mathbb{Z},$$

$$\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k), j \in \mathbb{Z}^+, k \in \mathbb{Z}$$

where α_k, β_{jk} are some coefficients of wavelet expansion and

$$\|f\|^2_{L_2(\mathbb{R})} = \sum_{k\in\mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} c_{jk}^2. \qquad (2.2)$$

There are several different ways of viewing the multivariate functional data. One of them is to apply wavelet decomposition on the data. Wavelets and PCA have also been combined by other researchers (see [19], [22], [23]), but the resulting techniques are different from the method proposed in this paper.

Let us give a definition of Sobolev space:

Definition 1. *Let $s \in \mathbb{N} \cup \{0\}$. The function $f \in L_2(\mathbb{R})$ belongs to the Sobolev space $W^s(\mathbb{R})$, if it is s-times weakly differentiable, and if $f^{(j)} \in L_2(\mathbb{R}), j = 1, 2, \ldots, s$. In particular, $W^0(\mathbb{R}) = L_2(\mathbb{R})$. Associate norm is*

$$\|f\|^2_{W^s(\mathbb{R})} = \|f\|^2_{L_2(\mathbb{R})} + \|f^{(s)}\|^2_{L_2(\mathbb{R})}. \qquad (1.1)$$

There is a strong relationship between norm in $W^s(\mathbb{R})$ and wavelet coefficients.

It was shown in [1],[4], that a function f lies in $W^s(\mathbb{R})$ if and only if

$$\sum_{k\in\mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} 2^{sj} c_{jk}^2 < +\infty. \tag{2.3}$$

Moreover, the discrete equivalent norm of Sobolev space $W^s(\mathbb{R})$ is

$$\|f\|^2_{W^s(\mathbb{R})} \approx \sum_{k\in\mathbb{Z}} c_k^2 + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} 2^{sj} c_{jk}^2, \tag{2.4}$$

where s is the smoothness order of the Sobolev space (see [20]). From this formula, it appears that switching from original data to wavelet expansion allows an efficient computation of the sobolev norm without having to explicitely estimate the velocity. Since we expect contaminating noise coming from wind for example, this permits some control on the noise part and the signal part.

Here and later we will use the equivalent norm. Let us construct an orthonormal basis in $H = (L_2(\mathbb{R}))^p$ using wavelets, where $p \in \mathbb{N}$. At first we must define an inner product and norm in the space H:

$$\langle \vec{f}, \vec{g}\rangle_H = \sum_{i=1}^{p} \omega_i^2 \langle f_i, g_i\rangle_{L_2(\mathbb{R})}, \tag{2.5}$$

$$\|\vec{f}\|^2_H = \sum_{i=1}^{p} \omega_i^2 \|f_i\|^2_{L_2(\mathbb{R})}, \tag{2.6}$$

where $\vec{f}(t) = (f_1(t), \ldots, f_p(t))^T, \vec{g}(t) = (g_1(t), \ldots, g_p(t))^T$ and $\omega_i,\ i = \overline{1,p}$ are positive weights. We use weights in order to avoid bigger influence from the bigger variability in one of the vector's component. The weights will be defined later.

If we let $\vec{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$, where "1" lies on the ith place, then the set of functions

$$\begin{aligned} &\{\vec{\varphi}_{ik}(t), \vec{\psi}_{ijk}(t)\}_{i=1}^{p}{}_{k=-\infty}^{+\infty}{}_{j=0}^{+\infty}, \\ &\quad \vec{\varphi}_{ik}(t) = \omega_i^{-1}\varphi_k(t)\cdot \vec{e}_i, \\ &\quad \vec{\psi}_{ijk}(t) = \omega_i^{-1}\psi_{jk}(t)\cdot \vec{e}_i \end{aligned} \tag{2.7}$$

forms an orthonormal basis in H which we will call a wavelet basis in $H = (L_2(\mathbb{R}))^p$ space.

As it was in case $p = 1$, any $\vec{f} \in H$ can be represented as a series:

$$\vec{f}(t) = \sum_{i=1}^{p}\sum_{k\in\mathbb{Z}} c_{ik}\vec{\varphi}_{ik}(t) + \sum_{i=1}^{p}\sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} c_{ijk}\vec{\psi}_{ijk}(x) \tag{2.8}$$

where c_{ik}, c_{ijk} are some coefficients, and

$$\|\vec{f}\|^2_H = \sum_{i=1}^{p}\sum_{k\in\mathbb{Z}} c_{ik}^2 + \sum_{i=1}^{p}\sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} c_{ijk}^2. \tag{2.9}$$

Moreover, if $\vec{f}$ lies in $(W^s(\mathbb{R}))^p$, then the discrete equivalent norm of Sobolev space is

$$\|\vec{f}\|^2_{(W^s(\mathbb{R}))^p} \approx \sum_{i=1}^{p}\sum_{k\in\mathbb{Z}} c_{ik}^2 + \sum_{i=1}^{p}\sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} 2^{sj} c_{ijk}^2, \tag{2.10}$$

IV. Functional Principal Component Analysis

A. Introduction

For many reasons, principal components analysis (PCA) of functional data is a key technique to consider (see [10], [11], [12]). Principal component analysis is among the most popular methods for extracting information from data, and has found application in a wide range of disciplines. PCA transforms the data in a statistically optimal manner by diagonalizing the covariance matrix by extracting the crosscorrelation or relationship between the variables in the data. If the measured variables are linearly related and are contaminated by errors, the first few components capture the relationship between the variables, and the remaining components represent the error. Thus, eliminating the less important components reduce the contribution of errors in the measured data and represents it in a compact manner.

Just as for the corresponding matrices in the classical multivariate case, the variance-covariance and correlation functions can be difficult to interpret, and do not always give a fully comprehensible presentation of the structure of the variability in the observed data directly. The same is true, of course, for variance-covariance and correlation matrices in classical multivariate analysis. A principal components analysis provides a way of looking at covariance structure that can be much more informative and can complement, or even replace altogether, a direct examination of the variance-covariance function.

B. Defining PCA for Hilbert space

The counterparts of variable values are vectors from the separable Hilbert space $\mathbb{H}$ (see [24]), which can be $L_2(\mathbb{R})$, $W^s(\mathbb{R})$ or more complicated space, so we will use terms "vector" or "function" in order to show the belonging to Hilbert space. When we were considering vectors, the appropriate way of combining a weight vector β with a data vector x was to calculate the inner product

$$\langle \beta, x\rangle = \sum_{j=1}^{p} \beta_j x_j. \tag{3.1}$$

When β and x are vectors from the separable Hilbert space $\mathbb{H}$, summations over j are replaced by inner product $\langle \beta, x\rangle$.

Within the principal components analysis, the vector $\beta \in \mathbb{R}^p$ now become a vector $\beta \in \mathbb{H}$. Using the notation, the principal component scores corresponding to weight β are now

$$f_i = \langle \beta, x_i\rangle. \tag{3.2}$$

In the first functional PCA step, the vector ξ_1 is chosen to maximize $N^{-1}\sum_{i=1}^{N} f_{i1}^2 = N^{-1}\sum_{i=1}^{N}\langle \xi_1, x_i\rangle^2$ subject to the analogue $\|\xi_1\|^2 = 1$ of the unit sum of squares constraint.

As for multivariate PCA, the weight function ξ_m is also required to satisfy the orthogonality constraint(s) $\langle \xi_k, \xi_m\rangle = 0, k \le m$ on subsequent steps. Each weight function has the task of defining the most important mode of variation in the curves subject to each mode being orthogonal to all modes defined on previous steps.

C. Multivariate PCA

For clarity of exposition, we discuss the extension of the PCA idea to deal with multivariate functional data. Suppose that the observed vector-functions are $\vec{f}^{(1)} = (x_1^{(1)}, \dots, x_p^{(1)})^T, \dots, \vec{f}^{(N)} = (x_1^{(N)}, \dots, x_p^{(N)})^T$, where $x_i \in \mathbb{H}$.

The most straightforward definition of an inner product between multivariate functions is:

$$\langle \vec{f}, \vec{g} \rangle_{\mathbb{H}^p} = \sum_{i=1}^{p} \langle x_i^{(f)}, x_i^{(g)} \rangle_{\mathbb{H}}. \tag{3.5}$$

The corresponding squared norm of a multivariate function is:

$$\| \overrightarrow{f} \|^2_{\mathbb{H}^p} = \langle \overrightarrow{f}, \overrightarrow{f} \rangle_{\mathbb{H}^p}. \tag{3.6}$$

What all this amounts to, in effect, is stringing more functions together to form a composite function. We now proceed exactly as in the univariate case, extracting solutions of the eigenequation system $V\vec{\xi} = \rho\vec{\xi}$, where $\vec{\xi}$ is a vector-function.

If the variability in one of the sets of curves is substantially greater than that in the other, then it is advisable to consider down-weighting the corresponding term in the inner product, and making the consequent changes in the remainder of the procedure

$$\langle \vec{f}, \vec{g} \rangle_{\mathbb{H}^p} = \sum_{i=1}^{p} \omega_i^2 \langle x_i^{(f)}, x_i^{(g)} \rangle_{\mathbb{H}}. \tag{3.7}$$

For example, the weights ω_i can be defined as

$$\omega_i = \hat{\sigma}_i^{-1} = \left(\frac{1}{N-1} \sum_{i=1}^{N} \| x_i - \frac{1}{N} \sum_{i=1}^{N} x_i \|^2_{\mathbb{H}} \right)^{-1/2} \tag{3.8}$$

One of the features of the functional data analysis approach to principal components analysis is that, once the inner product has been defined appropriately, principal components analysis looks formally the same, whether the data are the conventional vectors of multivariate analysis, scalar functions, or vector-valued functions. Indeed, principal component analysis for other possible forms of functional data can be constructed similarly; all that is needed is a suitable inner product, and in most contexts the definition of such an inner product will be a natural one.

Much of our subsequent discussion of PCA, and of other functional data analysis methods, will use univariate functions of a single variable as the standard example. This choice simplifies the exposition, but in most or all cases the methods generalize immediately to other forms of functional data, simply by substituting an appropriate definition of inner product.

V. Application to aircraft landing trajectories

A. Data collection

We now consider aircraft landing trajectories in Toulouse airport "Blagnac" (see "Fig. 3"). Without loss of generality for clusterization we can suppose that all aircraft landing trajectories are defined on a given interval $\mathcal{T}$ and have landed at the same time. Now suppose that we have a set of N curves. Each trajectory can be described as a curve in $\mathbb{R}^3$ space. Formally, we have observations which consist of triple $\vec{f}^{(i)} = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)})^T$, where x_1, x_2, x_3 are functions on interval $\mathcal{T}$. There is no way to measure $\vec{f}^{(i)}(t)$ on each point on interval $\mathcal{T}$, because aircraft trajectories are measured with radars. We have only access to samples at given time. A simple approach is to discretize the observed vector-functions $\vec{f}^{(i)}$ to a fine grid of n equally spaced values t_j that span the interval $\mathcal{T}$ (perhaps after a resampling). This yields a $N \times 3n$ data matrix

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3), \tag{4.1}$$

$$\mathbf{X}_i = \begin{pmatrix} x_i^{(1)}(t_1) & \dots & x_i^{(1)}(t_n) \\ \vdots & \ddots & \vdots \\ x_i^{(N)}(t_1) & \dots & x_i^{(N)}(t_n) \end{pmatrix}. \tag{4.2}$$

B. Wavelet filtering

In practical case the data matrix $\mathbf{X}$ is decomposed to its wavelet coefficients using the same orthonormal wavelet for each row into a matrix $\mathbf{XW}_3$, where $\mathbf{W}_3$ is an $(3n \times 3n)$ orthonormal matrix representing the orthonormal wavelet transformation operator containing the filter coefficients (see [22])

$$\mathbf{W}_3 = \begin{pmatrix} \omega_1 \mathbf{W} & 0 & 0 \\ 0 & \omega_2 \mathbf{W} & 0 \\ 0 & 0 & \omega_3 \mathbf{W} \end{pmatrix}, \tag{4.3}$$

and

$$\mathbf{W} = (\mathbf{G}_1, \dots, \mathbf{G}_L, \mathbf{H}_L), \tag{4.4}$$

where $\mathbf{G}_m$ is the matrix containing wavelet filter coefficients corresponding to scale m and $\mathbf{H}_L$ is the matrix of scaling function filter coefficients at the coarsest scale, which are coming from the usual orthonormal wavelet transformation operator in the space $L_2(\mathbb{R})$. The weights ω_i will be chosen as defined in (3.18).

Therefore, in case of the Sobolev space the data matrix $\mathbf{X}$ is decomposed to its weighted wavelet coefficients. It can be expressed as a transformation $\mathbf{X}$ into a matrix $\mathbf{XW}_3\mathbf{D}_3$,

$$\mathbf{D}_3 = \begin{pmatrix} \mathbf{D} & 0 & 0 \\ 0 & \mathbf{D} & 0 \\ 0 & 0 & \mathbf{D} \end{pmatrix}, \tag{4.4}$$

where $\mathbf{D}$ is a diagonal matrix defined by weights of wavelet coefficients in (2.20).

C. Applying MFPCA

In case of Sobolev space it is a very expensive deal to use a functional PCA. In order to avoid complicated calculations of functional PCA in Sobolev space we will use a PCA of weighted wavelet coefficients.

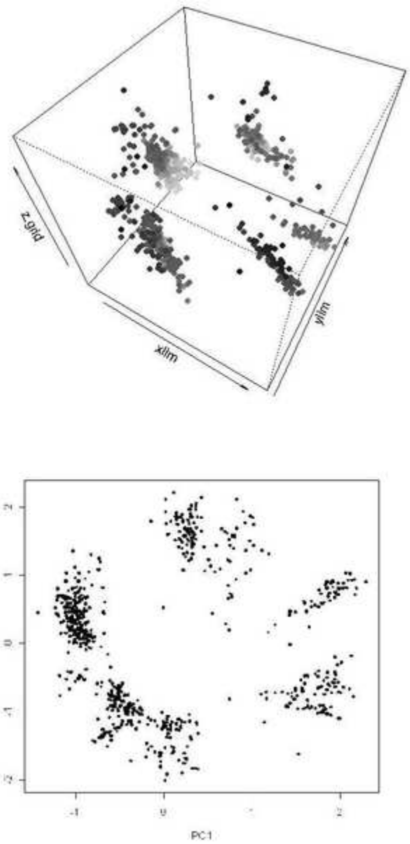

Figure 4. The scores of the aircraft trajectories on the three first principal components of $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ variation.

The covariance matrix of the weighted wavelet coefficients is

$$\mathbf{C} = (\mathbf{X}\mathbf{W}_3\mathbf{D}_3)(\mathbf{X}\mathbf{W}_3\mathbf{D}_3)^T = \mathbf{X}\mathbf{W}_3\mathbf{D}_3\mathbf{D}_3^T\mathbf{W}_3^T\mathbf{X}^T =$$
$$= \mathbf{X}\mathbf{W}_3\mathbf{D}_3^2\mathbf{W}_3^T\mathbf{X}^T = \sum_{i=1}^{3} \omega_i^2 \mathbf{X}_i \mathbf{W}\mathbf{D}^2\mathbf{W}^T\mathbf{X}_i^T. \tag{4.5}$$

For aircraft landing trajectories in Toulouse airport "Blagnac" three first principal components explain approximately 90% of the variability, but the third principal component does not play important role in clusterization (see "Fig. 4"). In-depth study of the role of x, y, z coordinates in clusterization problem gives us interesting conclusion: pairs (x, z) and (y, z) are pairs of independent random functions. Therefore, MFPCA of (x, y) coordinates must give "better" result instead MFPCA of (x, y, z) coordinates. The independence of z coordinate is a corollary fact of the air traffic controller action.

In case of MFPCA of (x, y) coordinates for aircraft landing trajectories the first principal component explain 75% of the variability, the second principal component explain 24% of the variability. In "Fig. 4" we can see the scores of the aircraft trajectories on the first two principal components of longitude-latitude variation.

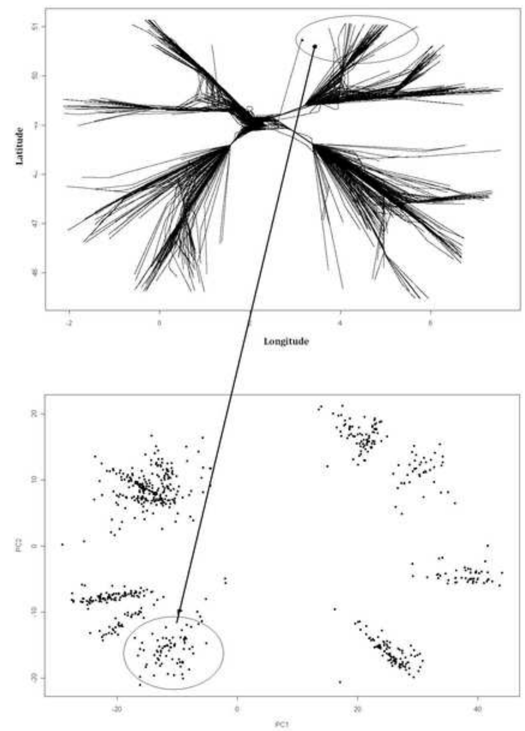

Figure 5. The scores of the aircraft trajectories on the first two principal components of longitude-latitude variation.

VI. Conclusion

We explored how multivariate functional principal component analysis in a Sobolev space using wavelets provides us spatial information which is important for certain clusterization task. Applying both wavelet decomposition and principal component analysis give efficient and practical way to cluster the data.

References

[1] I. Daubechies, Ten lecture on wavelets, *SIAM*, 1992.

[2] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inf. Th.*, 1990.

[3] C.K. Chui, An Introuduction to wavelets, *Xian Jiaotong Univ. Press*, 1994.

[4] Y. Meyer, Wavelets and Operators, *Cambridge Univ. Press*, 1992

[5] Y. Meyer, Wavelets: Algorithms and Applications, *SIAM*, 1993.

[6] W. Hardle, D. Picard, G. Kerkyacharian and A. Tsybakov, Wavelets, Approximation and Statistical Applications, *Un premier resultat du seminaire Paris-Berlin*, 1997.

[7] S.A. Dianat, R. Rao, Wavelet transforms: Theory and applications, *Proc. SPIE 12th Annu. Int. Symp. Aerosense*, 1998.

[8] M.A. Cody, The fast wavelet transform, *Dr. Dobb's J.*, 1992.

[9] Rong-Qing Jia, Jianzhong Wang and Ding-Xuan Zhou, Compactly supported wavelet bases for Sobolev spaces, *Appl. Comput. Harmon. Anal. 15*, 2003.

[10] I.T. Joliffe, Principal component analysis, *Springer-Verlag*, 1986.

[11] J.E. Jackson, A user's guide to principal components, *John Wiley*, 1991.

[12] J. Ramsay, B. Silverman, Functional Data Analysis, *Springer series in statistics*, 1997.

[13] J. Ramsay, B. Silverman, Functional Data Analysis (second edition), *Springer series in statistics*, 2005.

[14] F. Ferraty, Ph. Vieu, Nonparametric functional data analysis, *Springer series in statistics*, 2006.

[15] J. Ramsay, G. Hooker, S. Graves, Functional data analysis with R and MATLAB, *Springer series in statistics*, 2009.

[16] J. Ramsay, B. Silverman, Applied functional data analysis: Methods and case studies, *Springer series in statistics*, 2002.

[17] R. Adams, Sobolev spaces, *Academic press*, 1975.

[18] I. Gihman and A. Skorohod, Introduction to the theory of random processes, *"Nauka"*, 1965

[19] Maya R. Gupta, Nathaniel P. Jacobson, Wavelet principal component analysis and its application to hyperspectral images, *IEEE*, 2007.

[20] Feng Lu, Zhaoxia Yang and Yuesheng Li, Wavelets approach in choosing adaptive regularization parameter, *Lecture Notes in Computer Science*, 2007.

[21] R. A. Devore, Fast wavelet techniques for near-optimal image processing, *IEEE Military Communications Conference Record*, 1992.

[22] Bhavik R. Bakshi, Multiscale PCA with application to multivariate statistical process monitoring, *AIChE Journal*, 1998.

[23] S. Kaewpijit, J. Le Moigne, T. El-Ghazawi, Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis, *IEEE*, 2003

[24] F. A. Ocana, A. M. Aguilera, and M. J. Functional Principal Components Analysis by Choice of Norm, *Journal of Multivariate Analysis*, 1998